

REC'D 29 JUL 2003

WIPO

PCI

BEST AVAILABLE COPY



Kongeriget Danmark

Patent application No.: PA 2002 00888

Date of filing: 12 June 2002

Applicant:
(Name and address) EquTech ApS
Forskerparken Aarhus
Gustav Wieds Vej 10
DK-8000 Århus C
Denmark

Title: Lydkorrektion

IPC: -

This is to certify that the attached documents are exact copies of the above mentioned patent application as originally filed.



Patent- og Varemærkestyrelsen
Økonomi- og Erhvervsministeriet

04 July 2003

Åse Damm

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)



PATENT- OG VAREMÆRKESTYRELSEN

12 JUNI 2002

Modtaget

1

REALISTIC TARGET FUNCTIONS AND FEASIBLE GOALS FOR ROOM CORRECTION DESIGN

1.1 A brief summary of the pre-requisitive knowledge

We know now that sound in both time and frequency domains is influenced quite different according to the positions of source and receiver. Different transfer paths evoke different patterns of separable reflections as well as different patterns of excited modal frequencies. Operating with the revealed transfer functions, equalisation essentially means creating an exact inverse function, and thus being able to entirely remove the room acoustics by deconvolution (in this context de-reverberation is quite synonymous). For the reasons discussed earlier that scenario only works in infinitely small points in the room, - hence it is a mathematically beautiful technique but indeed not a feasible approach to practical system design.

Are anechoic surroundings preferable? Also, which sometimes is neglected, human beings do in fact not prefer sound reproduced in total anechoic surroundings. Some acoustic information must be present in to create a comfortable listening situation, and it is not sufficient to include that piece of information in the recording (neither as a live acoustic event nor as artificial reverberation). So, total de-reverberation (exact equalisation, inverse filtering) is not favourable from a qualitative point of view. It is relevant here to remember the inherent loudspeaker deficiencies namely, apart from non-ideal on-axis impulse characteristics, the usual lack of low frequency reproduction in the sub-octaves below app 50 Hz and the non-ideal and sometimes unsmooth off-axis characteristics.

1.2 Invertibility of room impulse responses

The Z-transform $H(z)$ of a measured room impulse response $h(n)$, although non-parameterised, can be ARMA modelled by a generalized digital IIR filter as in eq C1 1

$$H(z) = \frac{\sum_{j=0}^M b_j z^{-j}}{1 - \sum_{i=1}^N a_i z^{-i}} = \frac{\prod_{j=1}^{M_{in}} (1 - a_j z^{-1})}{\prod_{i=1}^{N_{in}} (1 - c_i z^{-1})} \frac{\prod_{j=1}^{M_{out}} (1 - b_j z)}{\prod_{i=1}^{N_{out}} (1 - d_i z)} \quad (C1 1)$$

Only stable systems of course are of interest leaving no room for any d_i . Systems with zeros inside the unit circle only (no b_j) are called *minimum phase* systems, and we refer to the phase contribution from the outside zeros represented by b_j as *excess phase*.

Transfer function decomposition

Through decomposition any transfer function $H(z)$ can be put into a product of a minimum phase part and an allpass part according to eq C1 2 with $H_{ap}(z)$ possibly containing also a pure delay. The minimum phase part consists of all the poles, the natural "inside" zeros, and any "outside" zero $z_{zero, excess}$ mapped to the inside with magnitude $1/r_{excess}$. The allpass part consists of the original "outside" zeros and poles cancelling out the artificially introduced zeros with magnitude $1/r_{excess}$. All possible magnitude information of $H(z)$ then is held in $H_{mph}(z)$, whereas the magnitude of $H_{ap}(z)$ as defined will always be unity.

$$H(z) = H_{mph}(z) H_{allpass}(z) \quad (C1 2)$$

We can invert $H_{mph}(z)$, but not $H_{allpass}(z)$ since the zeros outside the unit circle in $H_{ap}(z)$ turn into poles when inverted and then creating an unstable system. This means that we can compensate for the magnitude and minimum phase, but not for the excess phase.

Homomorphic deconvolution

Separation of minimum phase systems and allpass systems can be accomplished by employing homomorphic deconvolution. It can be shown that if a signal contains minimum phase only, then its cepstrum will turn out to be causal. Similarly, given a causal cepstrum, it is ensured that it represents a time domain signal containing minimum phase only. Consequently the minimum phase part of a response can be extracted by first forming the cepstrum, then deleting any non-causal information, and finally returning to the time domain. The cepstrum is formed by eq C1 3

$$\hat{h}(n) = IDFT_L \left(\ln |DFT_L \{h(n)\}| \right) \quad (C1 3)$$

The Hilbert transform

The minimum phase part is found in the cepstral domain by multiplying $\hat{C}[n]$ with $2u[n]-\delta[n]$, leaving only the causal part of the cepstrum. The all-pass part is then determined in the frequency domain by dividing the spectrum of the minimum phase part into the original spectrum. For large values of L these operations approximate the discrete Hilbert Transform which unambiguously links together magnitude and minimum phase, see [54]. Thus the minimum phase part $h_{\text{mph}}(n)$ of a response $h(n)$ can also be found by the magnitude of $H(z)$ since through the Hilbert transform the minimum phase can be derived from $|H(z)|$.

Non-causal excess phase equalisation Inverting a maximum phase system $h_{\text{max}}(n)$ leads to instability. However using the discrete time systems definition, see e.g. [54], it can be shown that the four combinations of the features *causality* and *stability* fall in two categories as in table C1.1

stable	causal		
stable	non-causal		
unstable	causal		
unstable	non-causal		

Table C1.1 Combinations of causality and stability

The interesting thing is that an unstable but causal system also can take the form of a stable but non-causal system, so by allowing non-causality the correction of maximum phase systems becomes possible. The excess phase in a room impulse response can then be equalised by introducing a delay. Ideally, when equalising $h_{\text{max}}(n)$ in a point-to-point scenario no artefacts are present in the correction delay part but the non-causal correction will introduce artefacts whenever the reproduction system is altered. The artefacts can be audible e.g. as pre-echoes and pre-reverberation. Presumably these audible phenomena produce a sound quality degradation, so in general excess phase equalisation is not recommended.

1.3 Recommendations for realistic and feasible goals

From the user's point of view, it must be considered realistic to require system operation based on only one initial microphone measurement in the optimised listening space. As literature shows, reasonable performance can be accomplished doing so, and when carefully designed perhaps only with minor drawbacks compared to a multi-microphone system. Also the system design should be aiming at stand-alone operation, thus pointing towards a fairly simple system not involving vast processing resources.

The more complex the better?

The room acoustical facts and the way human hearing works fortunately both speak against trying to build up a very complex correction system. At least in the sense of employing a very accurate correction scheme. Although it may seem tempting from a mathematical point of view, a very detailed correction would always only apply to very limited parts of the room which cannot be satisfying from a practical point of view. An optimised listening space of at least 1 m^3 must be required.

Is accuracy really positive?

Additionally, we tend to define the term *accurate* only from a technical viewpoint. Maybe such accurate correction does not correlate very well with perceived sound quality, and maybe we fool ourselves if striving for such accuracy. Accuracy may not be a positive goal in this case! One must be aware that sound quality (and sound quality improvement) will always be a rather diffuse and subjective measure, depending on a subtle and not fully understood combined time and frequency behaviour. The human hearing does not comply with the way technical equipment measures room acoustics and performs analyses on impulse responses. As accurate they well may be, in room correction design we are dealing with humans evaluating the improvements - and presumably not being able to discard personal preferences. The really tricky thing is to deal with 'non-accuracy' at an appropriate level! However, within that framework the following phenomena below must and can be dealt with. Notice that the phenomena are closely coupled to fig. A2.29.

1.4 Targets in the time domain

In the time domain it was shown to be appropriate to separate early from late parts of the impulse response using the statistical time t_{stat} .

Targets below the statistical time

In the early part separable reflections (or maybe the combined pattern of the first 5 to 8 reflections) should be considered. At least the most predominant, usually the first floor reflection, must be reduced in magnitude below the limit of audibility. Preferable the first 5-8 reflections should be reduced by 6-10 dB.

Targets beyond the statistical time

In the late part, statistically modelled, not much can be done. Reverberation time for average rooms is approx. 0.4 sec, and techniques should reduce RT if considerably larger than that in order to bring up the sensation of a more controlled listening room. Techniques for whitening the reverberation tail should also be applied so that no single frequency region is excessively represented, with due respect to the fact that the further out in the tail the less high frequency content should remain.

1.5 Targets in the frequency domain

In the frequency domain it seems appropriate to separate low frequencies from high frequencies putting the limit around the Schroeder frequency, f_{schr} , which for average listening rooms amounts to 100-200 Hz.

Targets below Schroeder frequency

In the low frequency region modal resonances are predominant in creating severe peaks and dips in the transfer function spectrum. Peaks which are audible as disturbing or even unpleasant resonances must be removed or reduced. As receiver and loudspeaker positions change so do the peaks and dips. Hence it may be hazardous to "fill up" the dips. When moving to other positions than the one equalised the dip compensation is probably less needed and a severe excess amplification is highly undesirable. For average listening rooms the bandwidth of even the narrowest modal resonance peak is 3-6 Hz, so compensation with a resolution of 2 Hz will be sufficient. Also, under f_{schr} the equaliser should aim at some energy compensation say in one third octave bands. The properties can be summarized in a desirable target band of $\pm 2\text{dB}$, but with quite large dynamics allowed regarded in the way that deviations from app 0 dB to -10 dB is allowed only in narrow bands.

Targets in the sub frequency region

Embedded in the measured room impulse response, the loudspeaker characteristics show off revealing little excitation of the room below app 50 Hz. As part of the correction, the low frequency reproduction should be extended down to app 25 Hz (or as far down as the loudspeaker is capable of handling the more power without introducing distortion). Below some f_{low} e.g. 25 Hz, there is no reason for further compensation. The human hearing is only little sensitive in this region and the equaliser might end up taking amplifiers and loudspeakers to the very edge of their performance due to most loudspeaker's natural roll off frequency no less than app 40 Hz. Hence the equaliser target also includes a lowpass filter.

Targets beyond Schroeder frequency

In the high frequency region not much can be done without introducing new unpleasant phenomena, but timbre should be considered i.e. the spectral energy should be equalised - presumably in no more detail than what can be done in one third octave bands. This very modest criterium complies well with the fact that considerable position sensitivity is present already at a few times f_{schr} and grows larger with frequency. Hence there is absolutely no physical reason for narrow band compensation at higher frequencies. Psychoacoustically it is difficult to detect a difference of 2 dB in two successive one third octave bands, so a reasonable target band is $\pm 1.5\text{ dB}$. Again we suggest a target roll off at high frequencies around 25 kHz, beyond which we may presume to have no interest in compensation. From app 1 kHz it may be beneficial to let the equaliser follow a slightly decaying target instead of a completely flat target, say 4-5 dB of total decay up to the upper limit of 25 kHz. Due to the larger absorption at high frequencies a room response will usually show a decay behaviour, and a subjective evaluation may prefer an equalised response that does not compensate for such introducing more high frequency energy.

Targets for phase characteristics

Although doubt still rules concerning audibility of transfer function phase, it must be recommended to strive for linear phase systems. Not an

easy task however when exact equalisation is not allowed. Smooth and not too excessive group delay will then be the second best goal.

Excess phase correction?

Another fundamental issue is: Can we ignore equalisation of the excess phase part in loudspeaker/room transfer functions? At the moment there is no clear answer, in an earlier investigation it has been shown that the excess phase, under certain circumstances is audible, see [36]. Equalisation of non-minimum transfer functions is generally problematic. We need a delay to obtain a causal impulse response, and therefore we can easily run into problems with pre-responses if we move the head to a position where the "compensation of the pre-response" is inaccurate, and such positions do exist. It is not clear how important it will be to separate the two parts of the compound impulse response for loudspeaker/room. Craven & Gerzon, [79], propose an equalization of the loudspeaker including non-minimum phase correction. In their opinion it is important to achieve a linear phase characteristic for the woofer highpass response. A recent review concerning equalization of loudspeakers is given by Karjalainen *et al.* in [74].

Equalising excess phase is a matter of accuracy. A high degree of excess phase correction is only possible when dealing with point-to-point transfer functions. These are generally not desirable, and as shown in fig. C1.1 and table C1.2 the transfer function excess phase increases with frequency and with time (the slight decrease beyond 300 ms is because the high frequency content reduces, see fig. C1.2). So fortunately the excess phase issue is not prominent in the regions where correction is feasible, and thus it will not be paid much attention henceforth.

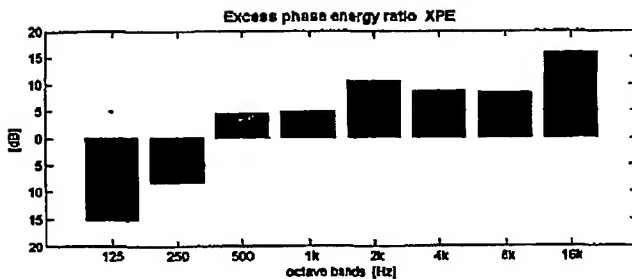


Figure C1.1 Excess phase as a function of frequency

section [ms]	XPE [dB]	section [ms]	XPE [dB]
0 - 75	-0.5	200 - 300	3.2
0 - 100	-0.4	250 - 350	6.4
50 - 150	-0.3	300 - 400	4.1
100 - 200	-0.4	350 - 450	2.6
150 - 250	0	400 - 500	3.9

Table C1.2 Excess phase as a function of time segment

1.6 Targets for energy relations and other acoustic parameters

Early/late energy relations

Controlling the time and frequency domains as suggested above usually also results in a smooth and non-transient behaving system regarding the energy relations measured by the room acoustical parameters like DR, C80, D50 etc. No direct action should be taken to improve in detail these parameters, however using them in objective evaluation is a powerful tool.

in a first hand judgement of the success of the correction. If one or more of the parameters show transient behaviour, it is most likely that subjective evaluation will reveal characterisations such as *annoying, disturbing, unpleasant*. Particularly Clarity and Direct-to-Reverberant signal energy relation though seems to play an important role for the perception of a high quality room. As a rule of thumb the experience tells us that Clarity (for small rooms C85 instead of C80) should exceed app 12 dB and DR should be 3-6 dB, see [88] and [90]

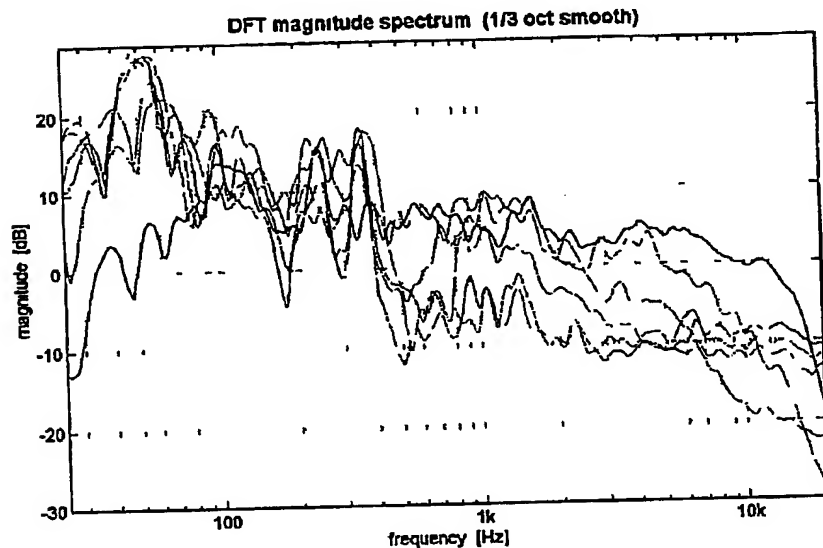


Figure C1.2 Spectral behaviour of the time segments in table C1.2. Black curve is based on the entire response length

Reverberation parameters

Instead of using a lot of effort to lower the reverberation time T_{60} (assuming a reasonably low initial value), it may be beneficial to simply go for a reduction in early decay time. That alone will contribute to the sense of a more damped room with a high subjective clarity.

Temporal repetition

Repetition of events (whole parts of the impulse response) is measured by the temporal diffusion Φ , and some diffusion technique may have to be considered if initially Φ is too small, i.e. below app 10 dB. Otherwise this parameter should just serve as a control measure.

1.7 Qualitative target specifications - IMOLE

Evaluating the equalised responses, the impulse response should be expected to possess as much "delta impulse" like behaviour as possible with no nasty long lasting resonances. Similarly, in the frequency magnitude no sudden jumps should be expected. Rather it shall look smooth with a slight decaying slope. It is also natural (but from a signal processing point of view not at all trivial) to introduce a criteria, call it IMOLE (IMprove Or LEave) saying that equalisers must be designed to generally

improve the sound reproduction or at least not deteriorate further the sound reproduction even in spaces away from the sweet spot set for correction

1.8 General issues to address in correction design

Which transfer functions to correct?

It will be assumed in the following that correction always is applied to the combined loudspeaker/room transfer function. Although by some decomposition techniques it could be possible to separate the effects of loudspeaker equalisation and room equalisation there are really no reasons for doing so - apart from mere curiosity. For very esoteric loudspeaker sets one can perhaps appreciate a pure room correction. A related issue concerns how to pre-process responses before the equalisation actions take place. Pre-smoothing or even averaging of more responses may serve to let the equaliser fulfill the subjective demands more easily.

Positioning of source and receiver

Just like the acoustic properties of the room in which we apply correction most definitely play a role for the final result, so will the initial (or assumed) positions of the loudspeakers and the listener. Different transmission paths correspond to different transfer functions these being input to a correction algorithm. From an intuitive point of view it is reasonable to assume that in order to reach to correction goals some transfer functions serve as more difficult inputs than others.

The preferable excitation

In a listening room with no electronic correction of sound reproduction when source and receiver positions fall together with many anti-nodes, we may experience annoying audible phenomena - the modal resonance sound becomes prevalent. On the other hand we completely lose low frequency reproduction if nodes of many resonances coincide with source and receiver positions. The best compromise being positioning of source and receiver where most strong and separable modal resonances are excited to some extent - let us say between 40% and 70%.

Positioning recommendation

From a correction point of view the more energy already present by inherent room resonance excitation the easier the correction algorithm design becomes. In the extreme case loudspeakers are placed in the corners and all modes and their combinations produce resonances. The correction algorithm can concentrate on reducing energy in order to meet the goals. Otherwise it may be forced to put in energy to make up for a possible set of poorly excited resonances. As long as the positions involved are totally fixed we can live with that scenario, but it is not hard to imagine what happens when for example the listener moves to another place in the room causing stronger excitation of the modal resonances.

2

THE ROOM CORRECTION DESIGN FRAMEWORK

2.1 Overview of the correction design framework

In fig C2 1 is shown a schematic of the framework built for loudspeaker/-room correction design. The main functions are preprocessing, band splitting, three band correction, summation, and post processing. The content of these building blocks are explained in detail in the following sections. The correction design system has been built up in a way to allow flexibility in all parameters. Although the design framework will correct a single response this may be composed by weighted averages of more responses. In the low frequency range where severe peaks occur a frequency resolution of app 2 Hz will suffice, see section 1. A direct implementation using an FIR filter requires around 22,000 filter coefficients to obtain 2 Hz resolution, and today this is still too heavy for standard signal processors. The high resolution is only required at low frequencies however so a band splitting and down sampling technique is obvious.

2.2 Pre-processing, band splitting, and resampling

In the first step an initial input response is derived from measured impulse responses. The initial response can be one single measurement or more impulse responses $h_i(n)$ may be averaged using arbitrary weights - within the entire bandwidth or if preferable just below some frequency f_{c_avg} . This allows for inputting a smoothed response to avoid or reduce position sensitivity at high frequencies or to implicitly make a better estimation of the perceived effects from low frequency resonances. A combination is also allowed, ie below f_{c_avg} the input response can be the average of responses from multiple sources to a single receiver position and beyond f_{c_avg} the single measurement rules. Still the point is to design a correction for one transmission channel at a time.

Band splitting and resampling

The initial input response is then split into three bands allowing for dedicated frequency dependent correction such as room acoustics and psychoacoustics point towards. The band splitting uses linear phase FIR filters in order to minimise any audible effects from these cross over filters. Four frequencies must be inputted. The low and high cut-off frequencies and the two crossover frequencies. It is reasonable to choose the lower crossover frequency in the neighbourhood of the Schroeder frequency of the room and the upper cross-over frequency 6-7 times higher where position sensitivity sets the agenda. For the high band the initial sampling rate is maintained but for reasons of convenience and the care for processing power the mid and low bands are resampled at rates 3-4 times the crossover frequencies.

BP filtered responses duration

In each band the duration of the response subject to equalisation can be set thus imposing a smoothing and reducing processing power. There are reasons to believe that the higher the frequency the shorter response is necessary.

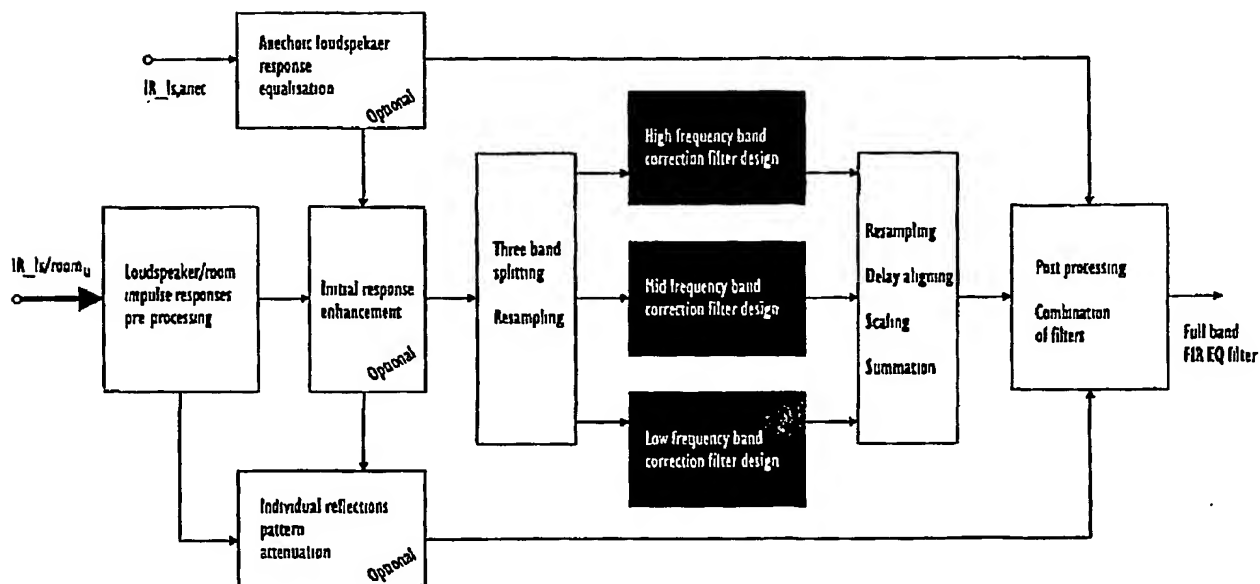


Figure C2.1 Overview of the correction algorithm design framework

2.3 Low frequency band correction filter design

The low-frequency channel is restricted to approximately the Schroeder frequency typically about 150 Hz indicating a sampling frequency below 1 kHz. In this case, a 2 Hz frequency resolution typically requires less than 500 taps. A robust inverse filter design method can be based on an AR model (all pole) of the input response. The inverse filter is based on

the LPC technique shortly described above and the order is variable. This compensation method is attractive because

- it particularly serves to suppress peaks,
- the equalising filter is all zero (MA), and stability is always ensured,
- the equalising filter is automatically minimum phase

Another way of creating an equalisation filter also incorporated is to simply invert the complex spectrum. Here however the spectrum subject to a regularisation before inversion in order to let the peaks weigh more than dips of the same magnitude. This method does not ensure minimum phase filters (only if the magnitude spectrum is used), and it tends to be inferior to the LPC method when it comes to robustness. Finally, together with any of the two magnitude related methods any amount of excess phase in the input response can be compensated for using a mirror convolution of the excess phase response - at the expense however of a delay equal to the length of the excess phase response.

2.4 Mid frequency band correction filter design

As described, the lower crossover frequency should be selected around the Schroeder frequency, and since position sensitivity is already a problem at a few times f_{chr} smoothing through a filter bank, with a resolution about 0.5 - 1 Bark could be motivated by psychoacoustics. In the frequency range above 500 Hz this resolution corresponds roughly to 1/6-1/3 octave. The Bark scale is more related to human sound perception (including timbre), and therefore it has been decided to investigate the performance of warped filters (WFIR), because they can be designed to approximate the Bark scale, see [75] and [72]. In the mid frequency band the following options are implemented

- AR modelling and inverse filter design by the LPC technique (or)
- minimum phase magnitude spectrum inversion
- pre-smoothing
- pre-warping
- reflections diffusion

The last option is a way of reducing the audibility of early strong reflections by convolving the response with a short (5 ms) exponentially weighted white noise response. This diffusion filter tends to blur the separable reflections somewhat but does no good for reverberation time and clarity. Again the AR model order is variable as are the smoothing factor (from 1 octave to 1/24 octave) and the warping factor allowing for putting more attention to the lower part of the mid band if enabled.

2.5 High frequency band correction filter design

In the high-frequency range the equalisation should preferably be reduced to correction of the tonal balance in eg 1/6 or 1/3 octaves. Note that the

psychoacoustically motivated Bark frequency scale is close to 1/3 octave, above 500 Hz. It is important to observe that the application of an FIR filter inherently includes a frequency smoothing caused by the window applied to limit the length of the filter response. In the high frequency band the following options are implemented

- minimum phase magnitude spectrum inversion
- pre-smoothing
- reflections diffusion

The reflections diffusion can be enabled in the high band too, and three alternatives of target functions are available. One with a flat frequency spectrum and two with slightly decaying spectra. The AR modelling method is not well suited for this band. It focuses on peaks but no narrow band equalisation is required (or even allowed) here. The functional blocks of the entire three band equaliser are shown in fig. C2.3

2.6 Auxiliary functions

To improve the correction performance two more options are included in the algorithm design framework. Both options (if enabled) alter the initial response to the three band equaliser, thus the three equalisation filters operate on the altered response, and the output of the three band equaliser must be corrected again. Going into the frequency domain and simplifying the three band equaliser to a blind inversion, the concept is shown by fig. C2.2. The input response $H(z)$ subject to correction design must end up with $1/H(z)$ regardless what happens on the way. The linear operation $R(z)$ of the auxiliary options must consequently be applied also after the inversion.

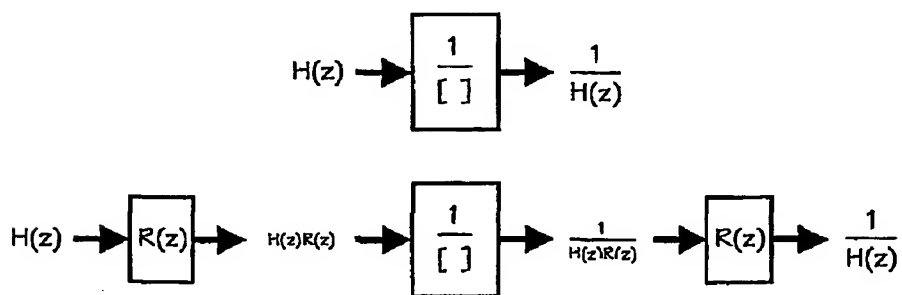


Figure C2.2 Pre- and post enhancement of the input response

Loudspeaker equalisation

For some reasons it may be advantageous to pre-equalise the loudspeaker and to include that equalisation filter in the algorithm operating on the entire input room response. Four ways of equalising the loudspeaker are proposed, see fig. C2.4

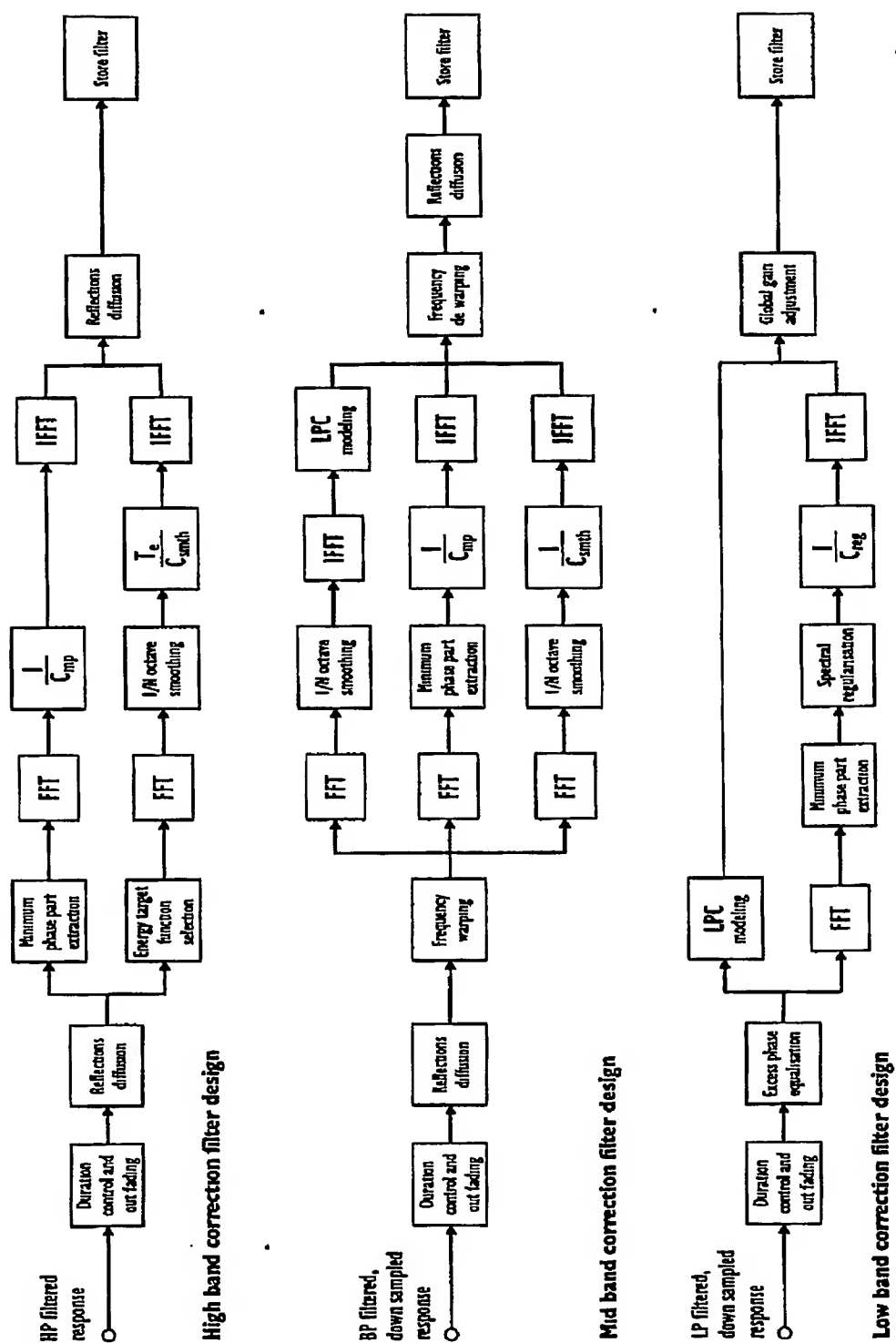


Figure C2.3 Functional blocks of the three band equaliser

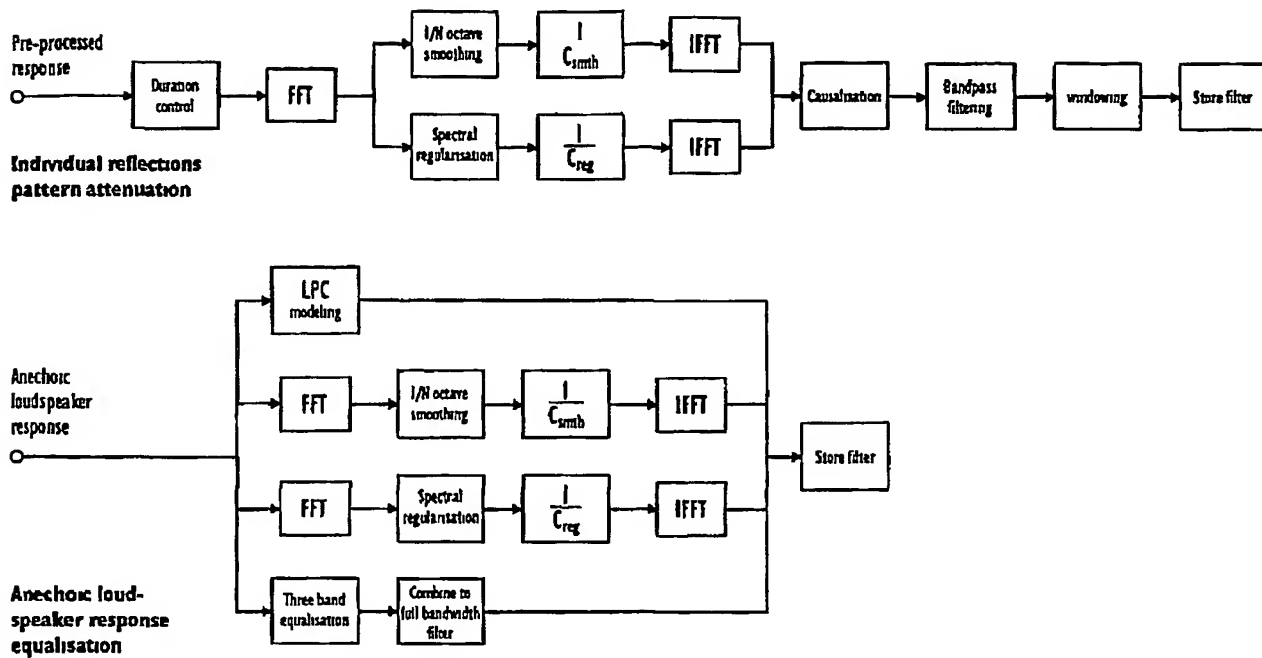


Figure C2.4 Reflections attenuation system and loudspeaker equaliser

Reflections attenuation

The three band equaliser mainly works in the frequency domain but to control the individual reflections in the input response it is necessary to operate in the time domain. The addressed reflections sequence is cut out, frequency transformed, and either subject to regularisation or smoothing before inversion to avoid a too sensitive correction of the reflections. By this modified deconvolution technique up to 30 ms of the response is attenuated by 6-12 dB by a reflections attenuation filter. It is not desirable to cancel out the reflections pattern entirely due to the position sensitivity issue and also because of the dubious subjective quality of a response with no energy at all in the first 15-30 ms. Both the regularisation and the smoothing call for a post causalisation, and finally the reflections attenuation filter is bandpass filtered to restrict its operation to the band 100-1000 Hz also to reduce the complete cancellation especially at high frequencies, see fig C2.4

2.7 Summation, post processing and operation of the system

After correction design in each band the correction filters are scaled and time aligned due to the possible delays introduced and finally put together into one FIR filter primarily for evaluations. A fade out window is applied and also for evaluation purposes the final filter is scaled in order to let a corrected response have the same energy (in the band 250 Hz to 5 kHz) as the initial response.

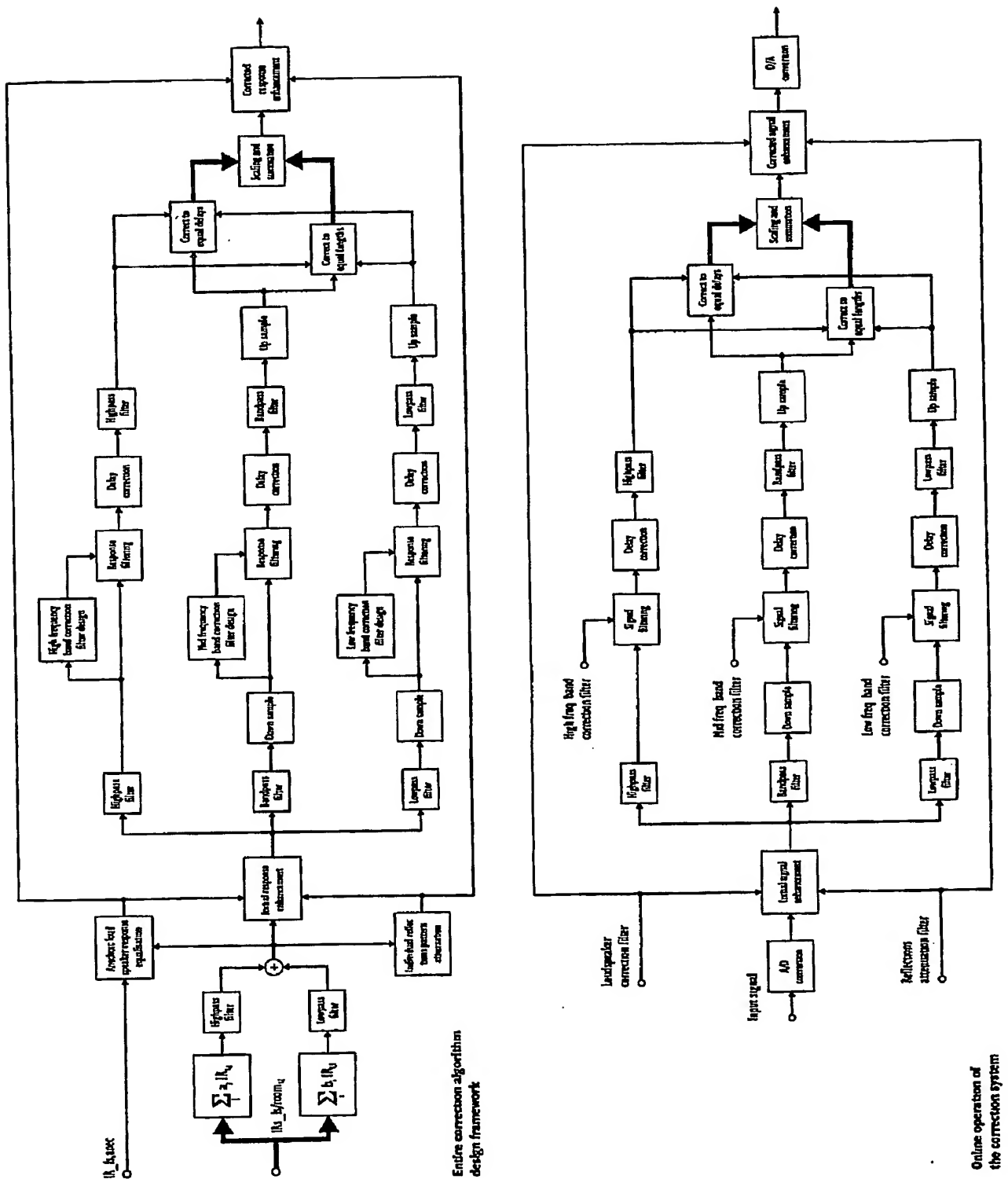


Figure C2.5 The two possible configurations of the correction framework, off-line for design of filters and on-line for signals correction

In fig C2 5 are shown the two possible configurations of the correction system, the off-line configuration where equalisation filters are designed based on measured responses and the on-line configuration in which electrical signals are corrected based on the stored equalisation filters

3

PARAMETER SETTINGS AND PRESENTATION OF CORRECTION ALGORITHMS FOR THE DALI ROOM

3.1 Initial correction algorithms design

The listening room at DALI Loudspeakers represents a set of very fine acoustic properties, in an absolute sense but indeed also compared to average listening rooms. The modal resonances are well distributed and the room is well damped due to wooden walls/ceiling and carpets on the floor and some of the walls. Even more measures have been taken to damp the room and hence only very few strong early reflections exist. Thus the energy distribution in impulse responses is pretty smooth and the acoustic temporal parameters are almost unbeatable, eg T_{60} and EDT close to 0.3 s. The initial algorithm designs aim at correcting the two-channel standard setup using source positions A and D and receiver position 0.

The challengeable preconditions

These properties pose a difficult starting point for any equaliser trying to improve sound reproduction. On the other hand, if the correction design framework is capable of producing algorithms that do in fact improve objective and subjective quality in this room, then one must suppose it could prevail in almost any room. This assumption makes the DALI room a difficult but interesting and challengeable room for evaluation really putting equalisers to the test.

Blind parameter settings

To start, all the flexible parameters described in section 2 have been varied and combined almost blindly producing approximately 75 different algorithms. Those include alternatives where receiver position smoothing is performed from weighted averaging of more impulse responses.

Pre-evaluation of the alternatives

All 75 alternatives have been pre-evaluated both in an objective sense through the impulse response analysis software described in part D and through informal listening test. Hence, the corrected impulse responses are subject to

- testing for temporal behaviour (early response, reverberation part), energy distribution, a few essential parameters (T_{60} , C35, EG), frequency magnitude behaviour, and group delay,
- listening events by the author - the response itself, and the response convolved with white noise and bandpass filtered pulses

The criteria for initial acceptance are

- temporal and spectral behaviour must be smooth and not plagued by local unexpected phenomena,
- acoustic parameters should be better than before correction (lower T_{60} / EG and higher C35),
- initial listening must not reveal any boomy, pumping, harsh, metallic, or strong reverberant behaviour,
- initial listening must comply intuitively with objective findings, ie no temporal or spectral band must be unintentionally emphasized

The pre-evaluation described above left back ten algorithms suitable for more thorough analysis. First thing to do was to evaluate the corrected impulse responses in all objective manners using the measures described in section A 2, all implemented in the analysis software. Secondly, all ten impulse responses were convolved with 12 different pieces of music to test the "naturalness" of the processing. Only one algorithm was picked out as the best compromise between objective and subjective performance but luckily, what performs best in the two senses seems to coincide.

Forming three interesting algorithms

From this basic "best" algorithm a further optimised version is formed, and two other versions are derived from curiosity. The first derivative dealing with the accuracy of low frequency phase equalisation and the second one dealing with explicit attenuation of strong reflections.

3.2 Band splitting and pre-processing facilities

Cross-over parameters

The cross-over frequencies of the three band equaliser were set to 150 Hz and 900 Hz respectively. The Schroeder frequency is app 95 Hz so above 150 Hz no individual resonance phenomena should be found, and the 900 Hz is chosen because of the mid frequency band corrections that are too delicate to be applied for higher frequencies. In fact any crossover frequency between 700 Hz and 1.5 kHz would suffice, however the cross-over of the particular algorithm selected as described above turned out to be 900 Hz. Lowest and highest correction frequencies are set to 25 Hz and 22 kHz respectively. Down sampling is performed to give new Nyquist frequencies at app 1.5 the cross-over frequencies (422 Hz and 2430 Hz respectively) which equal down sampling factors 144 and 25.

The cross-over filters are all linear phase FIR filters, and the orders have been chosen from the criterion that when adding down sampled bands of an ideal impulse with no corrections applied the result should come as close as possible to an unfiltered ideal impulse. Also, the slopes of LP and HP filters (for both crossover frequencies) should be approximately the same. This results in lowpass filter orders (taps) of 18, 28, and 18, and highpass filter order of 28, 84, and 560.

Pre-processing actions

The response input to the band splitting and down sampling is formed as the equally weighted sum of responses A0 and D0 below 150 Hz, and above 150 Hz no averaging is done. This averaging is introduced in order to better capture the general resonance phenomena instead of just the ones separately invoked by the loudspeaker positions A and D respectively. Slightly less accurate correction of the transfer functions A0 / D0 is the cost however. Finally the response is scaled till its total energy equals 1.

3.3 Digital signal processing and parameter settings

Correction features - low band

In the low frequency band it is chosen to determine an autoregressive (AR) model describing the transfer function. This model $1/A(z)$ consists of poles only and hence describes well the modal resonance peaks. The AR model is found by Linear Predictive Coding (LPC), and the number of coefficients is set to 48 resembling the effect of 24 second order poles. It is assumed (and verified) that 24 such poles should be sufficient to model the separable resonances up to 150 Hz. Using the $A(z)$ polynomial as FIR equalisation filter removes characteristic peaks in the transfer function without also undesirably putting energy into the natural dips in the transfer function. To compensate for the loss of energy to this peaks attenuation the entire low band is amplified 1.5 dB. In the low band equalisation operates on the whole input response i.e. 500 ms yielding an inherent smoothing of 2 Hz.

Correction features - mid band

In the mid band only the first 150 ms of the input response is used (frequency resolution of approx 7 Hz), and also here the AR modelling technique is applied. A first try suggested 144 coefficients producing fantastic objective results but listening tests revealed that there was a better way. Using the frequency pre-warping technique it becomes possible to focus more on low frequencies, and using a warping factor of 0.72 the LPC mathematics pays more attention to the band 150-400 Hz than to frequencies above 400 Hz. It is assumed that as frequency increases the transfer function phenomena easily modelled by AR poles also become less, i.e. there is good reason for combining AR modelling and pre-warping. Now the number of AR coefficients can be reduced to 48.

Correction features - high band

The high frequency band deals with the first 50 ms only yielding a frequency resolution of 20 Hz. In this band a straight spectrum inversion is

applied but prior to inversion the input response spectrum is smoothed in quarters of an octave. The smoothing removes any phase information, it is restored however (at least the minimum phase) using the Hilbert transform relations. After inversion the spectrum is weighted by a slightly decaying function (-4 dB from 1 kHz to 10 kHz) resembling the natural high frequency attenuation in room impulse responses, and finally transformed back to a time domain FIR filter.

Second phase evaluations

The algorithm with parameters as described above has now been tested of course in the objective sense where it performs well but also in a more realistic subjective sense. Different pieces of music have been pre-equalised with the algorithm, stored on a CD, and played back in the very listening room where the initial responses were recorded using of course the same loudspeakers and the same loudspeaker/listener positions. More listeners were invited to give their opinions revealing that the algorithm seems to be robust. The correction is found to be good in the sweet spot, which is approx 1 m², and outside the sweet spot the reproduced sound does not seem to be severely deteriorated.

3.4 Recommendation of three correction algorithms

In figs C3.1 and C3.2 the initial algorithm performance is shown. Grey plots show the input response and its spectrum and the black curves show the corrected response / spectrum. In the spectrum plots particularly it is easy to see the correction effort.

First algorithm derivative

To investigate the importance of low frequency phase correction accuracy the initial algorithm is slightly altered. First, the input response is not loudspeaker position averaged below the 150 Hz, and then the excess phase equalisation is applied as described in section C.2. Not for the entire response length (listening test showed this was not a good idea) but for 200 ms. The plots in figs C3.3 and C3.4 show the performance. In the time domain the performance is slightly better and the low frequency spectrum also looks nice apart from the transition to the mid band. The mid and high bands have been delayed appropriately corresponding to the low band delay caused by the excess phase equalisation.

Second algorithm derivative

In the second derivative of the algorithm the reflections attenuation capability is investigated. The input response is once again the low frequency position averaged one but now before the three band equaliser the reflections attenuation function is enabled. For the first 10 ms the reflections are set to be reduced (not totally removed) approx 8 dB, and that clearly shows on fig C3.5. Letting the reflections attenuated responses through the three band equaliser does not affect the resulting spectrum much, see fig C3.6. It still looks fine and pretty much as the one for the initial algorithm which is quite in accordance with expectations since the same algorithm parameters are used and the output response is post corrected with the reflections attenuation filter.

PART C

PARAMETER SETTINGS AND PRESENTATION OF CORRECTION ALGORITHMS FOR THE DALI ROOM

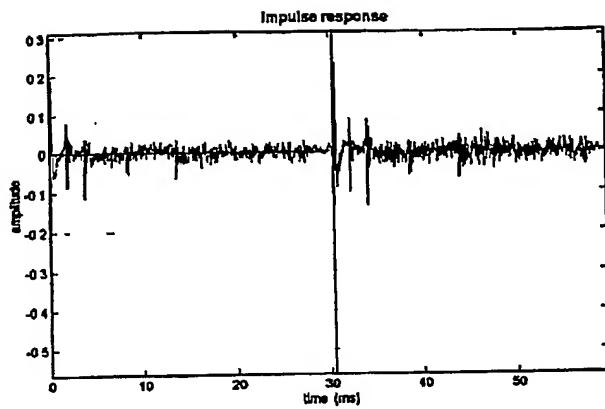


Figure C3 1 Initial algorithm behaviour in the time domain - black curve shows equalised response

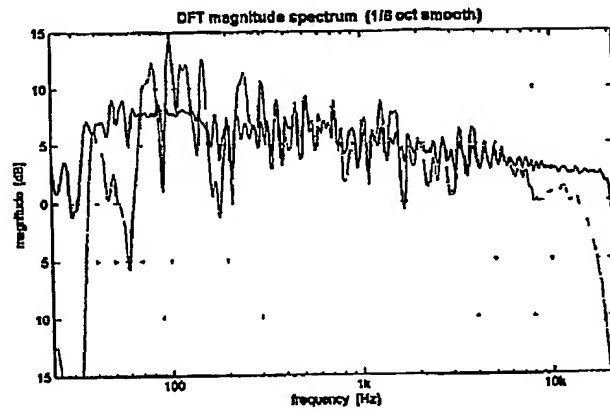


Figure C3 2 Initial algorithm behaviour in the frequency domain - black curve shows equalised response

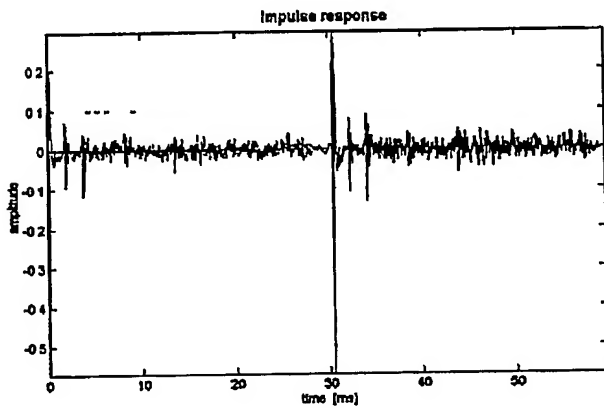


Figure C3 3 First derived algorithm behaviour in the time domain - black curve shows equalised response

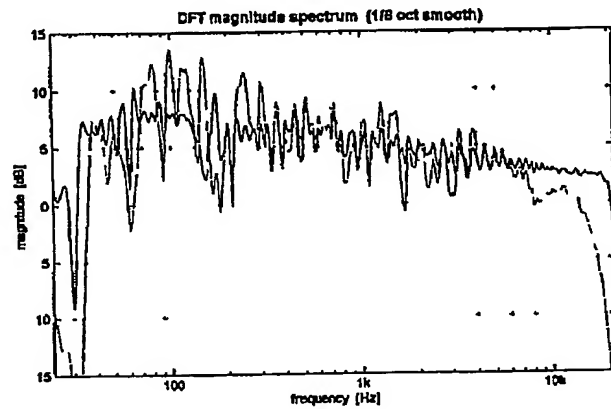


Figure C3 4 First derived algorithm behaviour in the frequency domain - black curve shows equalised response

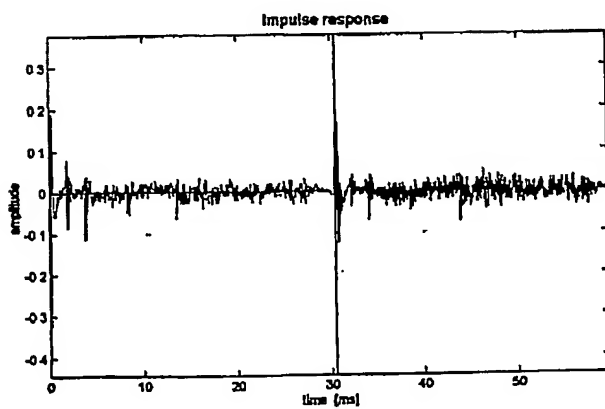


Figure C3 5 Second derived algorithm behaviour in the time domain - black curve shows equalised response

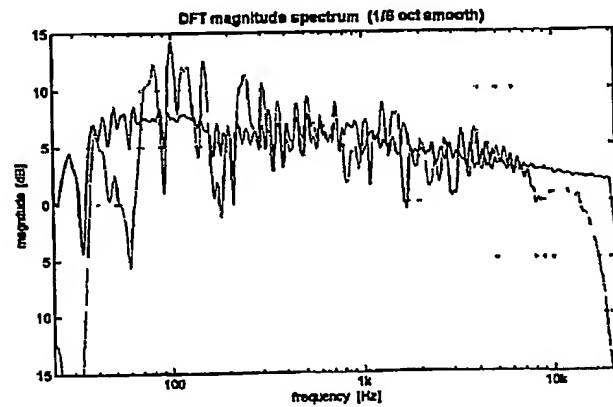


Figure C3 5 Second derived algorithm behaviour in the frequency domain - black curve shows equalised response

3.5 Presentation of three alternative correction algorithms

To test other features of the correction algorithm framework three alternative corrections have been designed. For each of these there is a specific purpose described below, and common for all three is that the subjective reproduction quality they represent has not been an issue.

Objective performance optimisation

The purpose of this algorithm is to show that whenever subjective performance is not an issue it is for sure possible to configure the design framework to come up with very accurate corrections. Actually, by a small listening test it has been verified that this algorithm in fact does not perform very good in that sense. No averaging is done for the input response, neither for listening positions nor for the loudspeaker positions at low frequencies. For all three bands the processed response length is 500 ms. In both the low and mid band AR modelling is applied, in the low band using 120 coefficients. In the mid band no smoothing and pre-warping is done, and as much as 288 LPC coefficients are imposed. Also in the high band smoothing and decaying function are omitted.

So from a signal processing point of view the actions taking place in the three bands more or less resembles that of a total spectral inversion due to the large number of LPC coefficients - only it happens in a minimum phase way. The spectral inversion is trivial apart from the excess phase, that is why the modelling technique tuned to higher accuracy is used. From other experiments it is well known that equalisers based on total spectral inversion corresponding to a blind deconvolution of the response does not correlate well with subjectively good performance. It simply becomes too accurate and position sensitive, but the objective performance is outstanding as shown in figs C3.7 and C3.8. Also the acoustic parameters turn out very well. In table C3.1 are put the characteristic acoustic parameters calculated for the input response as well as the corrected response. Particularly EDT, TD, and EG looks astonishingly good. Also D/R and the energy distribution through the Clarity numbers indicate a very accurate correction heading for a perfect impulse.

Acoustic parameters	before correct	after correct	Acoustic parameters	before correct	after correct
RT [s]	0.31	0.36	D50 [dB]	-0.4	-0.2
EDT [s]	0.29	0.16	SNR [dB]	2	8.8
RDL [%]	94	44	D/R [dB]	0.8	7.8
TS [ms]	43	38	TD [dB]	21.7	46.2
C20 [dB]	5.8	10.8	XPE [dB]	-0.4	-0.2
C50 [dB]	10.7	12.4	EG [%]	39	12
C80 [dB]	12.9	13.1	STI [0, 1]	0.76	0.74

Table C3.1 Room acoustic parameters before and after correction with optimised algorithm

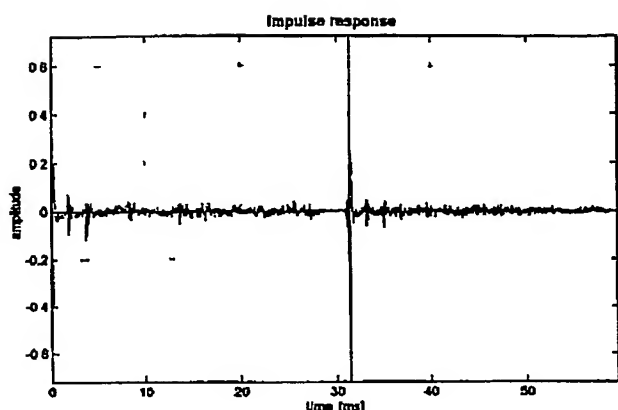


Figure C3 7 Theoretically optimised algorithm behaviour in the time domain - black curve shows equalised response

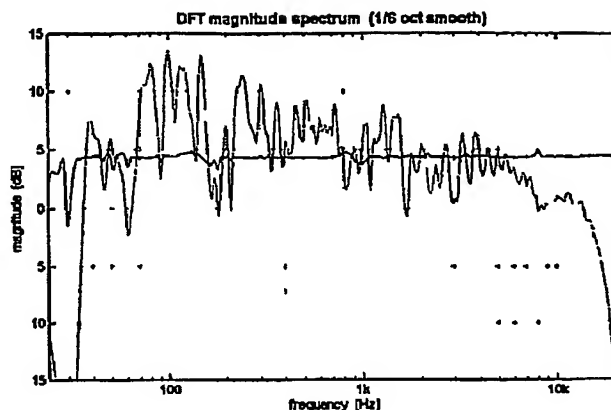


Figure C3 8 Theoretically optimised algorithm behaviour in the frequency domain - black curve shows equalised response

Reflections diffusion

Taking the same initial algorithm as described in section 3.4 the reflections diffusion feature has been enabled, as told in section 2.4 another way of rendering the first separable reflections inaudible. Instead of reducing their amplitude here they are blurred by a short exponentially decaying FIR filter of length 1.25 ms. It shows in figs C3 9 and C3 10 that the reflections are no longer visible as individual phenomena and that the diffusing feature does not deteriorate the frequency magnitude behaviour.

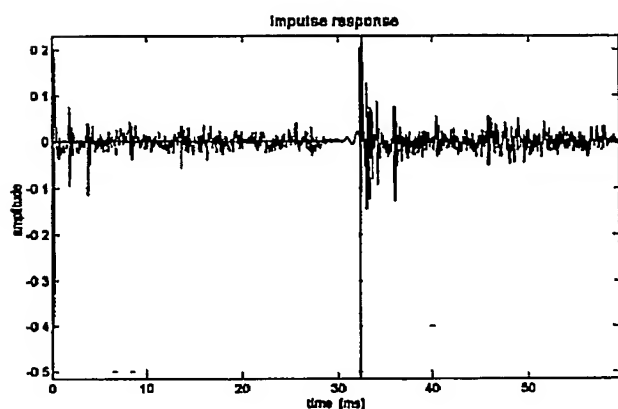


Figure C3 9 Correction with reflections diffusion functionality enabled - time domain

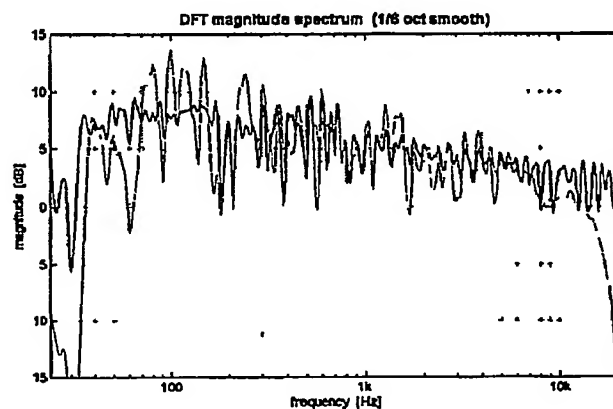


Figure C3 10 Correction with reflections diffusion functionality enabled - frequency domain

Surround setup corrections

In the last alternative also the same settings as in the initial algorithm have been used but now the correction is applied to five loudspeaker positions all to the same listening position as in a standard surround sound reproduction setup. That corresponds to measurements A0 / D0 (front speakers), I0 (middle speaker), and H0 / E0 (rear speakers). In figs C3 11 and C3 12 the five responses with the correct respective delays

have been added, and it is seen that after summation the corrections barely show. Also when adding responses the magnitude spectrum will still tend to decay app 3 dB per decade. In figs C3 13 and C3 14 the individual five corrected responses/spectra are shown. From top to bottom they come in the order H0, A0, I0, D0, E0. Actually, the added responses may say more about the perceived quality of the corrected reproduced sound, only it becomes difficult to analyse the effects of the correction in detail. As five channel surround reproduction equipment starts to take over from standard two channel stereo sets, correction electronics must be able to deal with five channel equalisation. It gives at least one advantage. Averaging the low frequency spectrum of five loudspeakers (or even six including a subwoofer) located around the room enables better capture of the general room resonances, and hence the risk that a resonance is not present in the measurements due to unlucky positioning of speakers/listener is minimised.

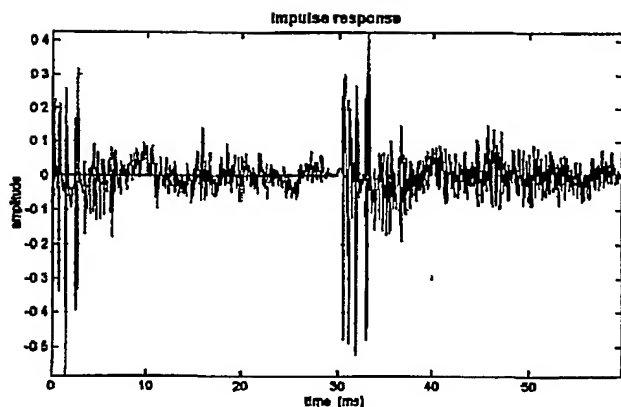


Figure C3 11 Summation of five corrected responses (black), H0, A0, I0, D0, E0 in the time domain

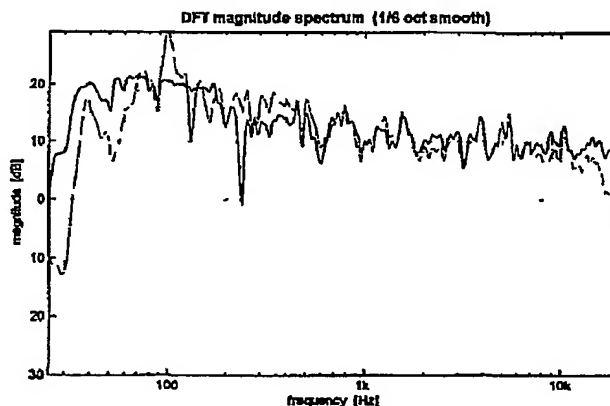


Figure C3 12 Summation of five corrected responses (black), H0, A0, I0, D0, E0 in the frequency domain

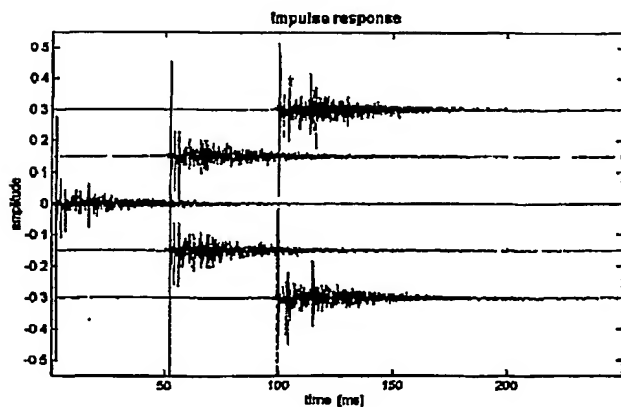


Figure C3 13 The five corrected responses, H0, A0, I0, D0, E0 individually in the time domain - delays are only for separation

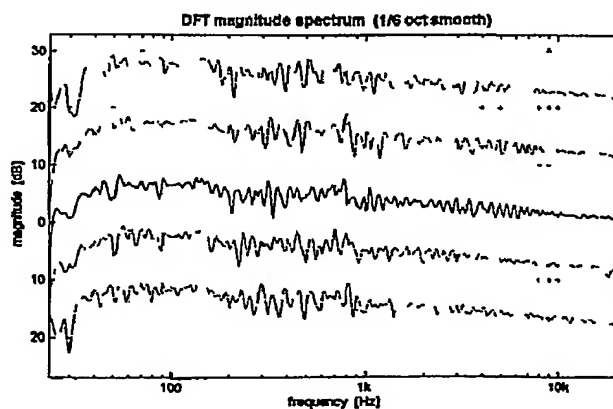


Figure C3 14 Five corrected responses, H0, A0, I0, D0, E0 individually in the frequency domain - only relative mag matters

3.6 Example of loudspeaker correction

To see what happens if loudspeaker equalisation only is desired, the anechoically measured active speaker has been subject to the same optimised parameters of the correction algorithm as were used in room correction alternative one. Figs. C3.15 and C3.16 show before/after responses and spectra, and Figs. C3.17 and C3.18 show the cumulative spectral decays before and after correction. The equalisation is quite prominent in all domains, but it is noticed that the price to pay is a very small pre-response of app. 0.5 ms.

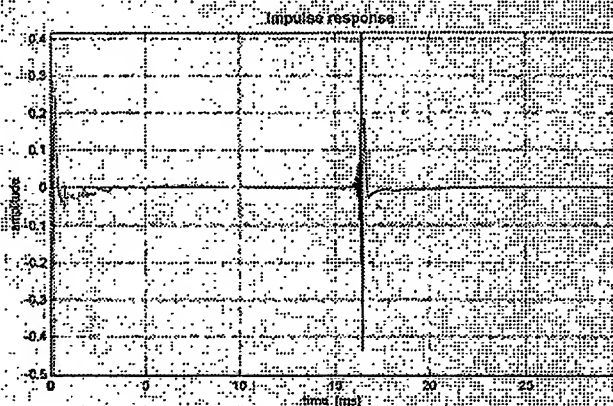


Figure C3.15: Loudspeaker equalised with optimised algorithm: time domain, black curve is corrected response.

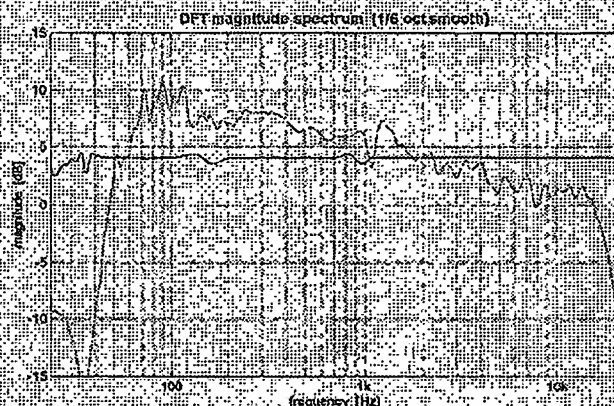


Figure C3.16: Loudspeaker equalised with optimised algorithm: frequency domain, black curve is corrected response.

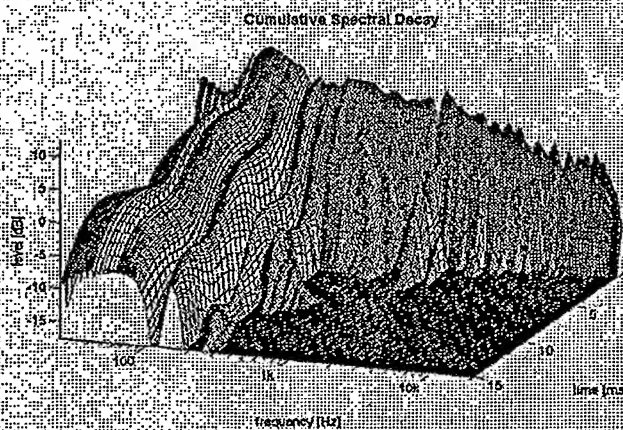


Figure C3.17: Loudspeaker Cumulative Spectral Decay before equalisation: using optimised algorithm.

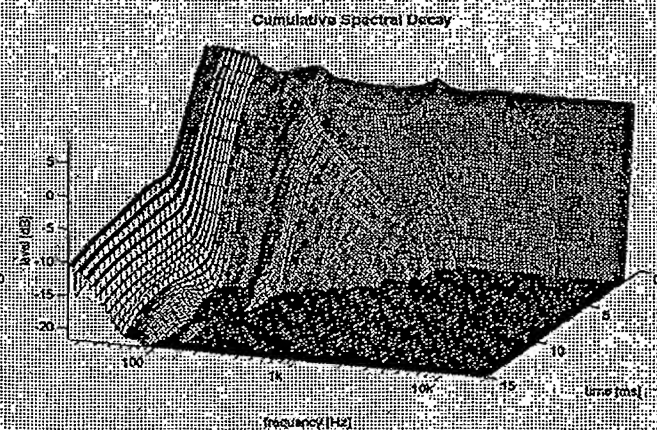


Figure C3.18: Loudspeaker Cumulative Spectral Decay after equalisation using optimised algorithm.

12 JUNI 2002

Patentkrav

Modtaget

- 1 Fremgangsmåde til kompensation for forvrængning af gengivet lyd fra højttalere i lokaliteter såsom beboelsesrum, kirker, banegårdspladser eller automobiler kendetegnet ved at kompensationen virker på et udvalgt område eller volumen uden negative konsekvenser for lyd kvaliteten udenfor det udvalgte område
- 2 Fremgangsmåde til kompensation for forvrængning af gengivet lyd fra højttalere i henhold til krav 1 kendetegnet ved at der i det udvalgte område korrigeres for faseforvrængning
- 3 Fremgangsmåde til kompensation for forvrængning af gengivet lyd fra højttalere i henhold til krav 1 eller 2 kendetegnet ved at lydreflektioner i det udvalgte område dæmpes
- 4 Fremgangsmåde til kompensation for forvrængning af gengivet lyd fra højttalere i henhold til et eller flere af kravene 1 - 3 kendetegnet ved at kompensationen foretages i 2 eller flere adskilte frekvensområder
- 5 Fremgangsmåde til kompensation for forvrængning af gengivet lyd fra højttalere i henhold til et eller flere af kravene 1 - 4 kendetegnet ved at korrektionen inkorporerer psykoakustiske forhold